PB90-142902

Determination of Polychlorinated Biphenyls Using Multiple Regression with Outlier Detection and Elimination

Wisconsin Univ.-Superior

Prepared for:

Environmental Research Lab.-Duluth, MN

1987

U.S. Department of Commerce
National Technical Information Service
NTIS

TECHNICAL REPORT DATA
(Please read Instructions on the reverse before comp

1. REPORT NO.	2.	PB90-142902
EPA/600/J-87/506		

4. TITLE AND SUBTITLE	5. REPORT DATE
Determination of Polychlorinated Biphenyls Using Multiple Regression with Outlier Detection and Elimination.	6. PERFORMING ORGANIZATION CODE
7. AUTHOR(S) Burkhard, L.P.[1] and D. Wininger[2]	8. PERFORMING ORGANIZATION REPORT NO.
9. PERFORMING ORGANIZATION NAME AND ADDRESS [1]Ctr. for Lake Superior Environ. Studies, Univ. of Wisconsin-Superior, Superior, WI 54880; [2] Chemistry Dept., Seaver Bldg., Pomona College, Claremont, CA 91711	10. PROGRAM ELEMENT NO.
	11. CONTRACT/GRANT NO. CR 812079
12. SPONSORING AGENCY NAME AND ADDRESS Environmental Research Laboratory Office of Research and Development U.S. Environmental Protection Agency Duluth, MN 55804	13. TYPE OF REPORT AND PERIOD COVERED
	14. SPONSORING AGENCY CODE EPA-600/03

15. SUPPLEMENTARY NOTES
Published In: Anal. Chem. 59:1187-1190 (1987).

16. ABSTRACT

A method for the analysis of capillary column polychlorinated biphenyl (PCB) data using regression analysis with outlier checking and elimination, COMSTAR, is presented and evaluated. This algorithm determines the best combination of the commercial PCB mixtures which best fits the chromatographic fingerprint of the sample by excluding weathered and contaminated PCB components from its final determination. Subsequently, significance testing on the final determination is performed. The extra sum of squares test is used for outlier testing. The chief advantage of COMSTAR over other PCB analysis method is its ability to discern more accurately the amount of PCB present in a sample when weathered and contaminated or enriched PCB components exist in the chromatographic data.

2 con't on pg 1187

17. KEY WORDS AND DOCUMENT ANALYSIS

a. DESCRIPTORS	b. IDENTIFIERS/OPEN ENDED TERMS	c. COSATI Field/Group

18. DISTRIBUTION STATEMENT	19. SECURITY CLASS *(This Report)*	21. NO. OF PAGES
Release to public	Unclassified	6
	20. SECURITY CLASS *(This page)* Unclassified	22. PRICE

EPA Form 2220-1 (Rev. 4-77)

REPRODUCED BY
U.S. DEPARTMENT OF COMMERCE
NATIONAL TECHNICAL INFORMATION SERVICE
SPRINGFIELD, VA. 22161

Anal. Chem. 1987, 59, 1187-1190 EPA/600/J-87/506

Determination of Polychlorinated Biphenyls Using Multiple Regression with Outlier Detection and Elimination

Lawrence P. Burkhard*

Center for Lake Superior Environmental Studies, University of Wisconsin—Superior, Superior, Wisconsin 54880

David Weininger

Chemistry Department, Seaver Building, Pomona College, Claremont, California 91711

A method for the analysis of capillary column polychlorinated biphenyl (PCB) data using regression analysis with outlier checking and elimination, COMSTAR, is presented and evaluated. This algorithm determines the best combination of the commercial PCB mixtures which best fits the chromatographic fingerprint of the sample by excluding weathered and contaminated PCB components from its final determination. Subsequently, significance testing on the final determination is performed. The extra sum of squares test is used for outlier testing. The chief advantage of COMSTAR over other PCB analysis method is its ability to discern more accurately the amount of PCB present in a sample when weathered and contaminated or enriched PCB components exist in the chromatographic data.

A number of recent papers (*1–5*) have presented methods for both quantitative and qualitative analysis of capillary gas chromatography data for polychlorinated biphenyls (PCBs) in environmental samples. We also have developed a regressive pattern matching algorithm, COMSTAR (complex mixture statistical reduction), for performing this type of analysis.

This laboratory as well as two others are currently using this algorithm for PCB determinations on environmental samples from a variety of sources. Among these laboratories, more than 2000 samples have been successfully analyzed in the past 4 years by using COMSTAR.

The difference between COMSTAR and the other methods is that an outlier detection and elimination algorithm for weathered and contaminated PCB components has been included in the COMSTAR method. With this additional algorithm, better fits of the PCB chromatographic fingerprint can be obtained since all peaks used in the final quantitative and qualitative PCB determination are members of a self-consistent PCB distribution.

In this report, a brief presentation and discussion of the COMSTAR method is given. In addition, a comparison between COMSTAR, SIMCA (*1*), regression with no outlier checking (*2–4*), and congener-specific (*6*) methods for quantitative and qualitative PCB determinations is presented for a few typical environmental samples.

ALGORITHM

COMSTAR is an iterative three-stage algorithm. Prior to entering stage one, analysis and output parameters must be set. These parameters include confidence levels, output display codes, and a toggle for weighted (default) or unweighted regression. In addition, the minimum number of PCB peaks present for a sample as well as the minimum number of largest PCB peaks in each Aroclor mixture must be specified.

Stage One, Regression Model. Stage one determines the combination of the Aroclor mixtures providing the best fit to the sample PCB distribution, assuming "perfect" input data (i.e., no PCB components are weathered or contaminated). Here, the sample PCB distribution is regressed against the Aroclor PCB distributions by using standard regression techniques (*7*). When performing this regression, COMSTAR treats a PCB peak absent from an Aroclor mixture as a null measurement (the quantity of the component is estimated to be zero). However, in a sample, absent PCB peaks are treated as missing data (absent peaks are not used in the analysis).

When performing the regression analysis, negative coefficients (concentrations) for the Aroclor mixtures may occur. Since negative matter cannot exist, stage one forces all coefficients to be nonnegative using the following iterative process. Here, if any of the coefficients are negative, the most negative coefficient is removed by fixing it at zero (temporarily, see stage two) and the regression is rerun. This process is repeated until all of the coefficients are nonnegative.

This method of constraining the coefficients was compared to a nonlinear least-squares method (*8*) using a linear model with boundary constraints, i.e., all coefficients ≥ 0.00. In nearly all cases, coefficients set to zero were predicted to be exactly zero by the nonlinear least-squares method. When nonzero coefficients were obtained, significance testing which occurs in stage three eliminated these coefficients. Since differences in the final COMSTAR solutions were insignificant between these two methods, the iterative method was selected due to its ease in implementation.

Stage Two, Outlier Detection. COMSTAR stage one results are the best-fit solution assuming "perfect" input data. Deviations caused by weathering or contamination of the PCB components may lead to erroneous stage one results. The function of stage two is to test data and exclude peaks which are not members of a consistent PCB population.

Outlying peaks are recognized as observations with high absolute standardized residuals (SR). Experience with environmental samples has shown that positive residuals are more common and cause more problems than negative residuals. Negative residuals (weathering) rarely exceed −20% and can never exceed −100% (none observed). Error introduced by positive residuals (contamination) has no natural bounds; e.g., DDE coelutes with a relatively minor PCB component and frequently causes residuals of +1000%. COMSTAR examines positive and negative residuals independently. Separate confidence limits can be specified for rejection of positive and negative outliers (α_1 and α_2, respectively). The worst outlying peak is defined as the observation with the highest positive SR as long as such observation fails the outlier test at α_1; otherwise the most negative SR is used.

In COMSTAR, the extra sum of squares test (Q_K) (*7, 9–11*) is used for outlier testing. Draper and John (*7, 11*) have recommended this method for outlier testing in regression analysis.

In performing the Q_K test, the worst outlying peak is temporarily removed from the data set and the regression is rerun.

0003-2700/87/0359-1187$01.50/0 © 1987 American Chemical Society.

NTIS is authorized to reproduce and sell this report. Permission for further reproduction must be obtained from the copyright owner.

The extra sum of squares statistic, a F-statistic, is

$$Q_K = (SS_{before} - SS_{after})/(SS_{after}/(n - p))$$

where the numerator is the difference of the residual sum of squares (SS) before and after removal of the test peak and the denominator is the variance of the regression after peak removal with n remaining peaks in a p predictor regression. The F-statistic (Q_K) is compared to the critical F-distribution value, $F(1,n-p,\alpha'_i)$. To ensure $1 - \alpha_i$ confidence since the test peak was not chosen randomly, α'_i is set to α_i divided by $n + 1$ (*11*).

Stage two test failure occurs when the F-statistic is greater than the critical F-distribution value. This means that we can be at least $1 - \alpha$ confident that inclusion of the test peak has a detrimental effect on the regression fit. Thus, the test datum is permanently removed from the data set and marked as an outlier. In this case, it is necessary to repeat stage one because the outlying observation which caused a stage two failure may have also been the cause of a negative result in stage one. Stage two elimination of an outlier causes all predictors previously removed in stage one to be reintroduced and stage one rerun. Eventually, the worst remaining negative outlier will pass the stage two outlier test. The test peak is then returned to the remaining data set and stage two terminates.

Stage Three, Testing Significance of Results. The final set of coefficients determined in stage two are reported by COMSTAR as the "best-fit result". In stage three, the "best-fit" coefficients (concentrations) for each Aroclor mixture are tested for significance (i.e., "Is each Aroclor mixture present at a concentration significantly greater than zero?"). Each coefficient is tested at the stage three confidence level, α_3, using the standard t-test for each coefficient (*7*). If one or more of the Aroclor mixtures are present at concentrations not significantly greater than zero, the least significant mixture is removed from the "best-fit result". The remaining Aroclor mixtures are then regressed against the same PCB components used in determining the best-fit result. Significance testing and Aroclor removal is repeated until all of the remaining components of the "best-fit result" are predicted to be present in significant quantities. COMSTAR outputs this as the "final result".

Output. Along with reports of results and confidences, COMSTAR provides a variety of optional post analysis displays. Overlaid printer plots of sample and reconstructed PCB chromatograms, the outlier elimination order, and residual plots are available.

EXPERIMENTAL SECTION

Reagents. Aroclors mixtures obtained from the EPA—Repository for Toxic and Hazardous Materials, US-EPA, EMSL, Cinninnati, OH, were used as the standard mixtures.

Samples. Environmental samples were taken and prepared for analysis as previously discussed (*12, 13*).

Equipment. A Hewlett-Packard (HP) Model 5710 gas chromatograph equipped with a Grob injector and a Ni-63 electron capture detector was used. Samples analyses were performed on a 60 m × 0.25 mm i.d. DB-5 capillary column (J+W Scientific) with a temperature program of 80–280 °C at 1 °C/min. A HP-3357 lab automation computer was used for data acquisition. Data was transferred via nine-track tape to a PDP 11/70 computer for COMSTAR analysis.

Program. A compiled–linked version of COMSTAR as well as the source code (in Fortran-77) are available at a minimal cost for an IBM-PC computer. Contact the senior author for further information if copies of the program are desired.

Procedure. A mixed standard consisting of Aroclors 1221, 1016, 1254, and 1262 with known amounts for each PCB congener was chromatographed. For congener-specific analyses, a retention time–congener specific response calibration table and, for peak area analyses, a retention time–response calibration table with response factors of 1.00 for all components were constructed on the lab automation system. Subsequently, the samples and standard Aroclor mixtures were chromatographed and analyzed by using the constructed calibration tables.

For COMSTAR analysis, integrated peak areas or congener amounts for identified PCB components and their retention times were used. The SIMCA analyses were performed as previously discussed (*1*). The congener-specific method above has been presented by Mullin (*6*).

RESULTS AND DISCUSSION

Four samples were analyzed by COMSTAR and these samples are typical of three commonly occurring PCB analysis situations; i.e., the PCB data from the chromatograph is "perfect" (no contamination or weathering), mildly "imperfect", or severely "imperfect". Results from these analyses as well as for five additional analyses are summarized in Table I. In addition, listings of the correlation coefficients of the regression (r^2) and the total predicted PCB concentration, ordered according to peak removal by the outlier detection and elimination algorithm, are displayed for three of these samples in Table II.

The first two samples in Table I, mixtures of Aroclor standards, illustrate the qualitative and quantitative behavior of COMSTAR with nearly "perfect" input data. Agreement between the nominal and predicted PCB concentrations and Aroclor distributions is excellent, i.e., less than 3% error. The behavior of the outlier detection and elimination algorithm with "perfect" input data is illustrated in Table II; a consistent PCB population is found by the elimination of a minimal number of peaks, and r^2 and total PCB values are nearly invariant.

The third sample (Table I), a Lake Ontario lake trout sample, illustrates the qualitative and quantitative behavior of COMSTAR with mildly "imperfect" input data. With all PCB peaks included in the quantification, COMSTAR predicted a total PCB concentration of 7.53 ppm with a r^2 of 0.226. The combination of stages one and two in COMSTAR seeks a subset of PCB peaks which forms a self-consistent PCB population. Here, COMSTAR eliminated four outliers before finding this subset of PCB peaks and the COMSTAR predicted PCB concentration was 7.05 ppm with a r^2 of 0.916. Examination of the r^2 and total predicted PCB concentration for this sample shows that r^2 increases toward 1.00 and the total predicted PCB concentration plateaus with increasing number of eliminated peaks (Table II). This is typical behavior of the outlier algorithm.

The fourth sample (Table I), a turtle sample (*13*), illustrates the behavior of COMSTAR with severely "imperfect" input data. For this sample, COMSTAR analysis failed, i.e., COMSTAR was unable to find a subset of PCB peaks which forms a self-consistent PCB population. This failure is shown by the smaller r^2 value, 0.725, and by the failure of r^2 to approach 1.00 with increasing number of eliminated peaks. COMSTAR results should not be reported for this sample. (We have listed the COMSTAR results for this sample for discussion purposes only.)

From our experiences with COMSTAR, acceptable COMSTAR solutions are obtained when the following conditions occur. First, r^2 for the analysis is greater than 0.90. Second, residuals for the regression are normally distributed with mean zero and contain no trends. Third, with increasing number of eliminated outliers, r^2 should approach 1.00. Fourth, with increasing number of eliminated outliers, plateauing of the total PCB concentration should occur. In addition, COMSTAR solutions should not vary with increasing α_1 and α_2 values.

We further recommend that plots of the sample and reconstructed PCB chromatograms (for the predicted compo-

Table I. Quantitative and Qualitative Analysis Results for PCBs

sample	nominal PCB concn	COMSTAR[a] r^2	outliers eliminated	total PCB	error,[c] %	Aroclor composition 1242	1248	1254	1260
mixture of aroclor standards	3.03[d]	1.00	5	3.03	0.0	0.34	0.33	0.32	0.01
mixture of aroclor standards	4.04[f]	1.00	5	4.08	1.0	0.24	0.27	0.24	0.25
Lake Ontario lake trout		0.92	4	7.05		0.00	0.27	0.47	0.26
turtle tissue, D (ID no. 10)[h]		0.72[i]	6	10.8		0.00	0.00	0.00	1.00
fish tissue, B (ID no. 12)[h]		0.90	3	3.07		0.00	0.26	0.24	0.50
toxaphene/PCB (0.28)[k]	4.04[f]	0.99	4	4.09	1.2	0.22	0.27	0.21	0.30
toxaphene/PCB (2.80)	4.04[f]	0.99	33	4.30	6.4	0.23	0.25	0.20	0.32
toxaphene/PCB (5.60)	4.04[f]	0.99	43	4.25	5.2	0.18	0.34	0.21	0.27
toxaphene/PCB (14.00)	4.04[f]	0.93[i]	19	24.2	499.0	0.00	0.00	0.00	1.00

regression with no outlier checking[b]						SIMCA				congener specific	
		Aroclor composition				Aroclor composition					
total PCB	error, %	1242	1248	1254	1260	1242	1248	1254	1260	total PCB	error, %
3.02	0.3	0.33	0.34	0.32	0.01	0.37	0.33	0.24	0.04[e]	2.97	2.0
4.05	0.2	0.25	0.26	0.23	0.26	0.17	0.24	0.29	0.28[e]	4.06	0.5
7.53		0.00	0.19	0.47	0.34		NA[g]			9.90	
21.3		0.01	0.00	0.18	0.81		RNDAA[j]			16.3[h]	
3.57		0.00	0.28	0.27	0.45		RADAA			3.2[h]	
4.21	4.2	0.24	0.24	0.22	0.31		NA			4.57	13.1
6.54	61.9	0.09	0.27	0.00	0.64		NA			10.2	152.5
11.3	179.7	0.09	0.14	0.00	0.77		NA			18.8	365.3
23.8	489.1	0.06	0.09	0.00	0.85		NA			37.5	828.2

[a] COMSTAR analysis conditions $\alpha_1 = \alpha_2 = 5.0\%$, $\alpha_3 = 1.0\%$, and weighted regression for the first five samples and $\alpha_1 = \alpha_2 = 20.0\%$, $\alpha_3 = 1.0\%$, and weighted regression for the remaining examples. [b] Analysis performed by using stage one of COMSTAR, only. [c] Error, percent difference between total and nominal PCB amounts. [d] Nominal composition, 1:1:1:0. [e] Different sample, ref 1. [f] Nominal composition, 1:1:1:1. [g] NA = not analyzed. [h] Reference 13. [i] Unacceptable COMSTAR analysis. [j] RNDAA = PCB residue could not be described by Aroclor or mixture of Aroclors. [k] Ratio of toxaphene to PCB in sample.

Table II. COMSTAR Total PCB Values and Correlation Coefficients of the Regression with Increasing Number of Eliminated Outliers[a]

mixture of Aroclor standards			Lake Ontario lake trout			turtle		
no. of eliminated outliers	r^2	total PCB	no. of eliminated outliers	r^2	total PCB	no. of eliminated outliers	r^2	total PCB
0	0.995	4.05	0	0.226	7.53	0	0.824	21.3
1	0.998	4.05	1	0.767	7.07	1	0.858	20.0
2	0.998	4.06	2	0.854	7.05	2	0.759	16.7
3	0.998	4.07	3	0.899	7.06	3	0.696	13.5
4	0.999	4.07	4	0.916	7.05	4	0.577	9.5
5	0.999	4.08				5	0.683	12.5
						6	0.725	10.8

[a] Samples in rows 2–4 in Table I.

sition) be made and evaluated when evaluating COMSTAR analyses. For mixed standards, the sample and reconstructed PCB chromatograms should be nearly identical. For environmental samples, these plots are typified by the Lake Ontario lake trout chromatograms displayed in Figure 1. These types of plots are extremely useful in reinforcing the credibility of the COMSTAR analyses.

When the above criteria for an acceptable COMSTAR analysis are not met, COMSTAR results should be used with *extreme* caution and probably should not be reported. In general, failure of a COMSTAR analysis indicates that the composition of the sample can not be adequately represented by a linear combination of the Aroclor mixtures.

To further evaluate the usefulness of COMSTAR, samples were analyzed by using COMSTAR and other PCB analysis methods, SIMCA (*1*), linear regression with no outlier checking (*2–4*), and congener-specific (*6*) methods. Results for these analyses are also summarized in Table I.

For mixtures of Aroclor standards, qualitative and quantitative results are in excellent agreement between all methods

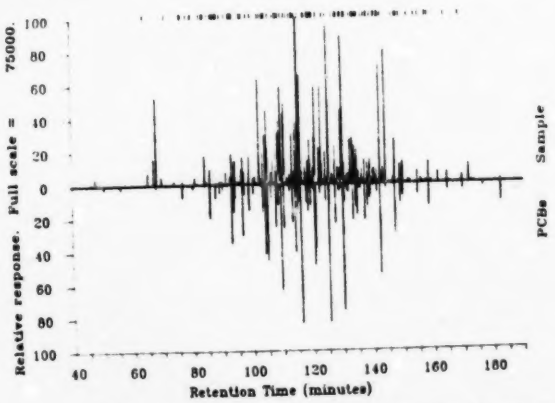

Figure 1. Display of Lake Ontario lake trout chromatographic data (above base line) and a PCB chromatogram, reconstructed from Aroclor standards and COMSTAR results (below base line). Tick marks above peaks indicate PCB congeners used by COMSTAR in final analysis.

(Table I). These results were anticipated since "perfect" input data were used.

For the turtle sample (Table I), an unacceptable COMSTAR analysis was obtained. Similarly, results obtained by using SIMCA, a principal component analysis technique (*1*), indicated that this PCB residue could not be described by an Aroclor or mixture of Aroclors (*13*). Analysis of this residue using regression with no outlier checking yielded a total PCB value of 21.3 and this value differs by ca. 30% from the total congener-specific PCB value reported by Schwartz et al. (*13*).

For the fish sample (*13*), differences between the various analysis methods did occur (Table I). For this sample, an acceptable COMSTAR analysis was obtained with the elimination of three outliers. In contrast, SIMCA analysis indicated that this PCB residue could not be described by an Aroclor or mixture of Aroclors (*13*). However, with the SIMCA analysis, all PCB components were used in making this decision. Comparison of the total PCB values determined by using COMSTAR and regression with no outlier checking to the congener-specific value reported by Schwartz et al. (*13*) (3.07, 3.57, and 3.2, respectively) illustrates the better quantitative ability of COMSTAR, i.e., differences of 4.1% and 11.6% from the congener-specific value, respectively.

The final samples analyzed in this comparison are a set of samples containing PCB contaminated with different amounts of toxaphene (Table I). For these samples with background contamination ranging from mild to severe, COMSTAR provided substantially better estimates of the total PCB values than regression with no outlier checking. For COMSTAR, errors of 1.2%, 6.4%, and 5.2% and, for regression with no outlier checking, errors of 4.2%, 61.9%, and 179.7% were obtained for toxaphene/PCB ratios of 0.28, 2.80, and 5.60, respectively. Additionally, qualitative information are more accurate for the COMSTAR analyses. For the toxaphene/PCB ratio of 14.0, COMSTAR analysis failed and yielded total PCB and composition values similar to that for regression with no outlier checking.

Errors in the qualitative description for these samples provided by COMSTAR are much larger than those observed in the total PCB values (acceptable analyses only). Larger errors in the qualitative description will occur in regression methods since the individual Aroclor mixtures are highly correlated. We, as thoroughly discussed by Draper and Smith (*7*), also deemphasize placing great importance on this qualitative information.

SIMCA analyses were not performed on the PCB samples contaminated with toxaphene. Since the PCB data from the gas chromatograph are composed of both PCB and toxaphene responses, we believe that SIMCA analyses would report that none of these samples could be described by an Aroclor or mixture of Aroclors. The major difference between COMSTAR and other analysis methods, i.e., SIMCA and regression with no outlier checking, is that COMSTAR assumes the input data is *not* pure PCB while the other methods assume that the input data is pure PCB. The outlier detection and elimination algorithm were included in COMSTAR attempts to unbiasly exclude the impure and weathered PCB components in calculating its best-fit result.

SUMMARY

A method for the analysis of capillary column PCB data using regression analysis with outlier checking and elimination, COMSTAR, has been presented and evaluated. The chief advantage of COMSTAR over the other PCB analysis methods is its ability to discern more accurately the amount of PCB present in a sample when numerous interfering chemicals are in the chromatographic data.

It is important that all analysts, when using COMSTAR as well as other computerized methods for the analysis of PCBs, understand the limitations of their methods. We recommend that "blind" use of COMSTAR be avoided.

We recommend four uses for COMSTAR. First, COMSTAR can be used as a quality control and assurance check for congener-specific analyses. Total PCB amounts determined using COMSTAR and congener-specific analyses should be similar provided acceptable COMSTAR solutions are obtained. Samples with large differences and/or with unacceptable COMSTAR solutions may be unusual and may require additional attention by the analyst. Second, COMSTAR can be used as a method for providing fast estimates of the total PCB values provided acceptable solutions are obtained. However, we do not recommend COMSTAR analyses if qualitative descriptions are desired. Other methods, SIMCA (*1*) and Burdick and Rayens (*14*), are more suited for this type of analysis. Third, COMSTAR analyses are useful in screening analyses especially when the actual PCB chromatograms are superimposed upon chromatograms constructed from the COMSTAR predictions. Examination of these plots will allow fast detection of samples which contain significant amounts of non-PCB components. Fourth, weathered and contaminated PCB components can be identified by using COMSTAR.

ACKNOWLEDGMENT

We thank Shelley Heintz for typing the manuscript, David L. Stalling for the fish and turtle data as well as the SIMCA analyses, Stephen J. Lozano for reviewing this manuscript, Sandra E. Beder-Miller for discussions on statistics, and David E. Armstrong for his support.

LITERATURE CITED

(1) Dunn, W. J., III; Stalling, D. L.; Schwartz, T. R.; Hogan, J. W.; Petty, J. D.; Johansson, E.; Wold, S. *Anal. Chem.* **1984**, *56*, 1308–1313.
(2) Capel, P. D.; Rapaport, R. A.; Eisenreich, S. J.; Looney, B. B. *Chemosphere* **1985**, *14*, 439–450.
(3) Liu, R. H.; Ramesh. S.; Liu, J. Y.; Kim, S. *Anal. Chem.* **1984**, *56*, 1808–1812.
(4) Schmitt, C. J.; Zajicek, J. L.; Ribick, M. A. *Arch. Environ. Contam. Toxicol.* **1985**, *14*, 225–260.
(5) Rayens, W. S. Ph.D. Thesis, Duke University, Durham, NC, 1986.
(6) Mullin, M. D. U.S. EPA Congener Specific PCB Analysis Workshop, June 12 and 13, 1985, Grosse Ile, MI.
(7) Draper, N. R.; Smith, H. *Applied Regression Analysis*, 2nd ed.; Wiley: New York, 1981.
(8) *SAS User's Guide: Statistics, Version 5 Edition*, SAS Institute: Cary, NC, 1985, pp 575–606.
(9) Gentleman, J. F.; Wilk, M. B. *Technometrics* **1975**, *17*, 1–14.
(10) John, J. A.; Draper, N. R. *Technometrics* **1978**, *20*, 69–78.
(11) Draper, N. R.; John, J. A. *Technometrics* **1981**, *23*, 21–26.
(12) Kuehl, D. W.; Johnson, K. L.; Butterworth, B. C.; Leonard, E. N.; Veith, G. D. *J. Great Lakes Res.* **1981**, *7*, 330–335.
(13) Schwartz, T. R.; Stallings, D. L. *Environ. Sci. Technol.* **1987**, *21*, 72–76.
(14) Burdick, D. S.; Rayens, W. S. *J. Chemometrics*, in press.

RECEIVED for review August 13, 1986. Resubmitted December 29, 1986. Accepted December 29, 1986. This work was supported in part by funding of the U.S. Environmental Protection Agency Cooperative Agreement CR-812079.

NOTICE

This document has been reviewed in accordance with U.S. Environmental Protection Agency policy and approved for publication. Mention of trade names or commercial products does not constitute endorsement or recommendation for use.

This article documents research that was conducted under the auspices of the U.S. Environmental Protection Agency (USEPA). As an arm of the government, the USEPA has a paid up, non-exclusive, irrevocable, worldwide license to reproduce, prepare derivative works, and distribute to the public all of the material contained herein, and to grant others acting on its behalf the right to do so. The USEPA grants to the National Technical Information Service, Department of Commerce, the right to act as its agent in the distribution of this material.